Instrumental Variable Estimation for Linear Panel Data Models

Lee, M.J.

Publication date:
1996

CentER for Economic Research

Discussion paper

[redacted]

Tilburg University

Center
for
Economic Research

No. 9641

INSTRUMENTAL VARIABLE ESTIMATION FOR LINEAR PANEL DATA MODELS

By Myoung-jae Lee

April 1996

ISSN 0924-7815

INSTRUMENTAL VARIABLE ESTIMATION FOR LINEAR PANEL DATA MODELS

(April 20, 1996)

Myoung-jae Lee

Department of Econometrics

Tilburg University

P.O.Box 90153

5000 LE Tilburg

The Netherlands

E-mail: M.J.LEE@KUB.NL

FAX: 31-13-466-3280

Key Words: panel data, instrumental variable estimator, method-of-moments.

INSTRUMENTAL VARIABLE ESTIMATION FOR

LINEAR PANEL DATA MODELS

We review instrumental variable estimators for linear panel data models where zero unconditional moment conditions hold for products of instruments and error terms across time; no conditional moment conditions nor high-order moment conditions are imposed. The models we consider do not specify the error term distributions and allow fixed-effect (correlation between the unit-specific error term and regressors), nonstationarity (different distributions across time), simultaneous panel data models (endogenous regressors), dynamic panel data models (lagged dependent variables as regressors), and time-varying parameters.

1. Introduction.

Consider a $\nu\times 1$ vector moment condition

$$E\rho(\omega,\beta)=0 \tag{1.1}$$

where $\omega\equiv(y,x')'$ is a random vector, and β is a $k\times 1$ $(k\le\nu)$ parameter vector of interest. An estimator based upon (1.1) is called a method-of-moments estimator (MME). For a cross-section linear model $y=x'\beta+u$, often we use $\rho(\omega,\beta)=u\cdot x=(y-x'\beta)\cdot x$; if x and u are correlated, then one can use $E(u\psi)=0$ where ψ is an instrument vector of dimension k or greater. An estimator based upon orthogonality between the error term and instruments is called an instrumental variable estimator (IVE), a special case of MME.

The main questions in MME are finding appropriate moment conditions, and then combining (more than enough) moment conditions to obtain an efficient estimator; for IVE, the first question is finding instruments. MME (so IVE) is easy to compute and its asymptotic properties are by now well known (e.g. Newey and McFadden (1994) and Lee (1996)); its applications are too numerous to mention.

As for efficiency of MME, if the (vector) moment condition can "explain" fully the score function of the maximum likelihood estimator (MLE) which requires specifying the distribution of ω, then the MME is as efficient as MLE; otherwise MME is less efficient. If the only available information is (1.1), then the generalized method-of-moments (GMM, Hansen (1982)) obtained by minimizing

$$(1/\surd N)\Sigma_i\rho(\omega_i,b)'\cdot\{V[(1/\surd N)\Sigma_i\rho(\omega_i,\beta)]\}^{-1}\cdot(1/\surd N)\Sigma_i\rho(\omega_i,b)$$

with respect to (wrt) b is the efficient way of using the moment condition (Chamberlain (1987)), where $V[\cdot]$ is the variance of $[\cdot]$ and $\{V[\cdot]\}^{-1}$ should be replaced by a consistent estimate in practice.

If we have the conditional moment condition

$$\text{(1.2)} \qquad E\{\rho(\omega,\beta)|x\} = 0,$$

then there is an estimator (Chamberlain (1987)), say estimator C, more efficient than GMM and as efficient as when the functional form of

$$\text{(1.3)} \qquad \sigma(x)^2 \equiv E\{\rho(\omega,\beta)\rho(\omega,\beta)'|x\} \text{ is known.}$$

For the linear model, $\sigma(x)^2=E(u^2xx'|x)=V(u|x)\cdot xx'$; if the form of $V(u|x)$ is known, then one can do weighted least squares (WLS) estimation. The estimator C (as in Robinson (1987)) is as efficient as the WLS estimator despite using only (1.2).

In this paper, we review IVE for linear panel data models using only unconditional first-order moment conditions as (1.1), but not conditional ones as (1.2) nor second-order ones as (1.3). Thus the models we consider do not specify the error term distributions, nor restrict the form of heteroskedasticity or the error term serial correlations. As in IVE for cross-section models, there are two issues for panel IVE: finding instruments and getting an efficient IVE. The second issue is solved by GMM, and hence the first is the main concern.

Suppose we observe N subjects (or units, individuals, firms) over T periods. Assume that N is large and T is small so that we can apply asymptotic theory wrt N but not wrt T. We take the following panel data model as our basis: for i=1...N and t=1...T,

$$\text{(1.4)} \qquad \underset{1\times 1}{y_{it}} = \tau_t + \underset{1\times k_c}{c_i'}\alpha + \underset{1\times k_x}{x_{it}'}\beta + \delta_i + u_{it} \equiv \tau_t + \underset{1\times k}{w_{it}'}\gamma + v_{it},$$

τ_t: unobservable, time-variant, unit-invariant,

c_i: observable, time-invariant, unit-variant,

x_{it}: observable, time-variant, unit-variant,

δ_i: unobservable, time-invariant, unit-variant,

u_{it}: unobservable, time-variant, unit-variant,

$k \equiv k_c + k_x$, $\gamma \equiv (\alpha', \beta')'$, $w_{it} \equiv (c_i', x_{it}')'$, $v_{it} \equiv \delta_i + u_{it}$;

the expressions "$\cdot \times \cdot$" below the variables denote the dimensions. We observe $(y_{it}, w_{it}')'$ and estimate $(\tau_1 \ldots \tau_T, \gamma')'$. Assume iid observations across i and allow arbitrary dependence across t within a given i; also allowed are nonstationary distributions across t. δ_i may be correlated with some or all of w_{it}; then (1.4) is called a "fixed-effect" model which is more general than "random effect" models where δ_i is constraint to be uncorrelated with w_{it}. Without loss of generality, set

(1.5) $$E(\delta_i)=0 \quad \text{and} \quad E(u_{it})=0 \quad \text{for all } t,$$

because both $E(\delta_i)$ and $E(u_{it})$ can be absorbed into τ_t.

An example of (1.4) is

(1.6) y_{it}: annual wage of married men of age 40 to 60,

τ_t: effect of the economy on y_{it} common to all i,

c_i: race, schooling years,

x_{it}: working hours, local unemployment rate, self-employment dummy,

δ_i: ability, IQ, or productivity,

u_{it}: unobserved time-variants, say, some variables for residential surroundings and individual characteristics.

In the classification of variables into five categories in (1.4), the category of time-variant and unit-invariant observables is missing. For example, a macro policy variable, say π_t, affecting all units in the same way may be relevant. But so long as π_t has a coefficient non-varying across i, it can be absorbed into τ_t.

There are a number of different ways the (composite) error term v_{it}

and the regressor w_{it} are orthogonal across time. The simplest one is $E(w_{it}v_{it})=0$ analogous to the usual cross-section moment condition. But also possible is $E(w_{is}v_{it})=0$ for all s,t. Thus, differently from cross-section data, there are in general more orthogonality conditions than the number of parameters in panel data, which then leads naturally to IVE and GMM for random-effect as well as for fixed-effect models.

For a random-effect model, the main issue is designing an efficient estimator using the available moment conditions and the error term structure $v_{it}=\delta_i+u_{it}$. These are also useful for fixed-effect models. For instance, given $E(v_{it}w_{it})\neq 0$, if $E(\delta_i w_{it})\neq 0$ but $E(u_{is}w_{it})=0$ for all s,t, then we can use

$$(1.7) \qquad E(\Delta v_{it}\cdot w_{it}) = E(\Delta u_{it}\cdot w_{it}) = 0$$

for the fixed-effect model where $\Delta v_{it}\equiv v_{it}-v_{i,t-1}$. Alternatively, if $w_{it}=\hat{w}_i+\tilde{w}_{it}$ with $E(\hat{w}_i\delta_i)\neq 0$ but $E(\tilde{w}_{it}\delta_i)=0$ for all t, then

$$(1.8) \qquad E(v_{it}\Delta w_{it}) = E(v_{it}\Delta\tilde{w}_{it}) = 0$$

is a valid moment condition. Both (1.7) and (1.8) take advantage of the moment conditions and the error term structure $v_{it}=\delta_i+u_{it}$.

(1.7) and (1.8) give an easy way of classifying the fixed-effect literature into two: one is "error-differencing (-transforming)" (1.7), and the other is "regressor-differencing (-transforming)" (1.8) where error terms appear intact in the moment conditions while regressors are transformed to yield instruments. The latter, which may be called the genuine IVE, is our focus. In some cases, the two approaches may give the same estimator, and one may be said to be the "dual" to the other.

The rest of this paper is organized as follows. In Section 2, we rewrite our basic model using compact notations and present four types of basic moment conditions. In Section 3, we present LSE, IVE and GMM for the moment conditions of Section 2. In Section 4, error-differencing ideas for fixed-effect models are reviewed; also there, we briefly review the "linear projection" idea of Chamberlain (1982,1984) for fixed-effect models. In Section 5, through an example, various instrumental variable matrices for different moment conditions are illustrated. In Section 6, two special variables (age and experience) which require some attention in practice are examined. In Section 7, regressor-differencing ideas for fixed-effect models are reviewed and an efficiency gain issue in using more instruments is discussed. In Section 8, a minimum distance estimation (MDE, see e.g., Lee (1992) and Lee (1996) and the references therein) is introduced, with which we can test the constancy of the model parameters across time; for the case the constancy is rejected, we extend (1.4) to time-varying parameter models. Finally, Section 9 concludes. Section 7 on efficiency gain and Section 8 appear to be new in the literature.

Throughout the paper, we will often use a T=3 case as an example. As for notation, we use $=^p$ for convergence in probability, and $=^d$ for convergence in distribution. If $\sqrt{N}(b_N-\beta)=^d N(0,\Omega)$ for an estimator b_N for β and $\Omega_N=^p\Omega$, then we will denote this simply as $\sqrt{N}(b_N-\beta)=^d N(0,\Omega_N)$. Zero vectors are denoted as 0. Assume that all random variables have the fourth moments; the second moments are enough for getting asymptotic distributions, and the fourth moments are sufficient to estimate the variance matrix, say $\Omega=E\{\theta_i(\beta)\}$, consistently with its sample analog $\Omega_N=(1/N)\Sigma_{i=1}^{N}\theta_i(b_N)$. Denote the correlation for λ_1 and λ_2 as $COR(\lambda_1,\lambda_2)$.

2. Basic Model and Moment Conditions for Panel Data.

Stack the model (1.4) for the unit i across t:

$$(2.1)\quad \begin{bmatrix} y_{i1} \\ \cdots \\ y_{iT} \end{bmatrix} = \begin{bmatrix} \tau_1 \\ \cdot\cdot \\ \tau_T \end{bmatrix} + \begin{bmatrix} c_i' \\ \cdots \\ c_i' \end{bmatrix}\cdot\alpha + \begin{bmatrix} x_{i1}' \\ \cdots\cdot \\ x_{iT}' \end{bmatrix}\cdot\beta + \begin{bmatrix} \delta_i \\ \cdot\cdot \\ \delta_i \end{bmatrix} + \begin{bmatrix} u_{i1} \\ \cdots \\ u_{iT} \end{bmatrix}.$$

This can be compactly written as (now c_i includes 1 and one of τ_t's drops out due to this; see the next paragraph)

$$(2.2)\quad \underset{T\times 1}{y_i} = \underset{T\times(T-1)}{m}\cdot\tau + \underset{T\times k_c}{(1_T'\otimes c_i)'}\alpha + \underset{T\times k_x}{x_i'}\beta + 1_T\delta_i + \underset{T\times 1}{u_i},$$

$$\equiv m\cdot\tau + \underset{T\times k}{w_i'}\gamma + v_i = \underset{T\times(k+T-1)}{q_i'}\cdot\eta + v_i,$$

where 1_T is the $T\times1$ vector of ones,

$$\tau\equiv(\tau_2,\ldots,\tau_T)', \qquad q_i'\equiv(m,w_i'), \qquad \eta\equiv(\tau',\gamma')', \qquad v_i\equiv 1_T\delta_i+u_i,$$

and m is a $T\times(T-1)$ time dummy matrix; with T=3, m is

$$(2.3)\quad m = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Estimation of τ_t requires time dummies. To see this, observe the following two equivalent ways of rewriting $(\tau_1\ldots\tau_T)'$ of (2.1) when T=3:

$$(2.4)\quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} = I_3\tau, \qquad \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} \tau_2-\tau_1 \\ \tau_3-\tau_1 \\ \tau_1 \end{bmatrix} \equiv (m,1_3)\tau^*$$

where $\tau^*\equiv(\tau_2-\tau_1,\tau_3-\tau_1,\tau_1)'$. Having $I_3\tau$ as in (2.1) means that we use three time dummy variables (I_3) to estimate τ. Alternatively we can use T-1 time dummies along with 1 included in c_i as in the second equation of (2.4); then the intercept estimate in α is τ_1 and the slope estimates of the T-1 time dummies are the deviations of τ_2 and τ_3 from τ_1.

One may ask what happens if we include 1 in c_i but no time dummies, perhaps because all τ_t's are believed to be the same. Depending on the model moment conditions, however, not using the time dummies can result in $(1/N)\Sigma_i v_{it} \neq 0$ for some t (i.e, not using the moment condition $E(v_{it})=0$ for some t); so it is advisable to use time dummies for the same reason as we use 1 as a variable in cross-section studies. From now on we will use T-1 time dummies and use 1 as the first element of c_i.

To estimate γ, we need moment conditions between the observable q_{it} and the unobservable v_{it}. For random-effect, there are a number of possible moment conditions: for all s and t indexing time periods,

(2.5) (SUM) (a) $E(\Sigma_t x_{it} v_{it}) = 0$ (b) $E(c_i \Sigma_t v_{it}) = 0$;

(CON) (a) $E(x_{it} v_{it}) = 0 \ \forall \ t$ (b) $E(c_i v_{it}) = 0 \ \forall \ t$;

(PRE) (a) $E(x_{is} v_{it}) = 0 \ \forall \ s \leq t$ (b) $E(c_i v_{it}) = 0 \ \forall \ t$;

(EXO) (a) $E(x_{is} v_{it}) = 0 \ \forall \ s, t$ (b) $E(c_i v_{it}) = 0 \ \forall \ t$;

where SUM, CON, PRE, and EXO respectively stands for Summation, Contemporaneous, Predetermined, and (strictly) Exogenous. Note that (b) follows from (a) if x_{it} were c_i. For fixed-effect, we need to replace either v_{it} by a transformed error term or x_{it} by instruments (recall (1.7) and (1.8)) while adjusting s and t somewhat; how this is to be done specifically will be shown later.

The following implication arrows hold:

(2.6) $$\text{SUM} \Leftarrow \text{CON} \Leftarrow \text{PRE} \Leftarrow \text{EXO}.$$

In SUM, x_{it} and v_{it} are allowed to be correlated. SUM is the moment condition for the LSE treating the panel data as N·T many cross-section data, because the kind of the moment condition used for the LSE is

$$(1/NT)\Sigma_i \Sigma_t x_{it} v_{it} =^p 0 \ \Leftrightarrow \ (1/N)\Sigma_i (1/T)\Sigma_t x_{it} v_{it} =^p 0 \ \Leftrightarrow \ E(\Sigma_t x_{it} v_{it})=0.$$

In CON, only contemporaneous correlations are zero. In PRE, x_{is} is allowed to be correlated with v_{it} if s>t (e.g., rational expectation models). In EXO, x_{is} and v_{it} are uncorrelated at all leads and lags. By omitting certain time periods in PRE and EXO, we can easily allow endogenous regressors.

3. LSE, IVE and GMM.

In a SUM type moment condition

$$(3.1) \qquad \underset{(k+T-1)\times T}{E(q_i} \cdot \underset{T\times 1}{v_i)} = E(\underset{(k+T-1)\times 1}{\Sigma_t q_{it}} \cdot \underset{1\times 1}{v_{it}}) = 0,$$

the number of moment conditions is the same as the number of parameters. In this case, the following LSE is the most efficient (analogously to cross-section studies):

$$(3.2) \qquad h_{1se} \equiv (\Sigma_i q_i q_i')^{-1} \cdot \Sigma_i q_i y_i .$$

It is easy to show that

$$(3.3) \quad \sqrt{N}(h_{1se}-\eta) =^d N(0,\ (\Sigma_i q_i q_i'/N)^{-1} \cdot (\Sigma_i q_i v_i v_i' q_i'/N) \cdot (\Sigma_i q_i q_i'/N)^{-1}).$$

With T=3, the moment conditions due to 1 in c_i and two time dummies are

$$(3.4) \qquad E(v_{i2})=0, \quad E(v_{i3})=0, \quad E(v_{i1}+v_{i2}+v_{i3})=0,$$

which is equivalent to $E(v_{it})=0$ for all t. If we do not include time dummies in q_i, we will be using only the third condition.

Turning to the other conditions CON, PRE and EXO, we need to use IVE as mentioned in Section 1, for there are more moments than parameters. Suppose we have an instrument matrix z_i such that $E(z_i v_i)=0$, where the column dimension of z_i is T and its row dimension is at least as large as that of q_i; we will show how to get z_i in practice through an example in Section 5. The IVE is

(3.5) $$h_{ive} = \{\Sigma_i q_i z_i'(\Sigma_i z_i z_i')^{-1}\Sigma_i z_i q_i'\}^{-1}\cdot\Sigma_i q_i z_i'(\Sigma_i z_i z_i')^{-1}\Sigma_i z_i y_i.$$

The GMM estimator is obtained by

(3.6) $$h_{gmm} = (\Sigma_i q_i z_i'\cdot C_N^{-1}\cdot\Sigma_i z_i q_i')^{-1}\Sigma_i q_i z_i'\cdot C_N^{-1}\cdot\Sigma_i z_i y_i$$

where $C_N \equiv (1/N)\Sigma_i z_i \hat{v}_i \hat{v}_i' z_i'$, $\hat{v}_i \equiv y_i - q_i' h_{ive}$, and

(3.7) $$\sqrt{N}(h_{gmm}-\eta) =^d N(0,\ \{(\Sigma_i q_i z_i'/N)\cdot C_N^{-1}\cdot(\Sigma_i q_i z_i'/N)\}^{-1}).$$

The IVE and GMM include the LSE as a special case when $z_i = q_i$. In the following, we show the specific form of z for CON, PRE and EXO.

Using $(T\cdot k)\times 1$ CON moment conditions requires setting

(3.8) $$z_i \equiv \mathrm{diag}(w_{i1},\ldots,w_{iT}),$$

for, with T=3,

(3.9) $$z_i v_i = \underset{(3\cdot k)\times 3}{\begin{bmatrix} w_{i1} & 0 & 0 \\ 0 & w_{i2} & 0 \\ 0 & 0 & w_{i3} \end{bmatrix}} \cdot \underset{3\times 1}{\begin{bmatrix} v_{i1} \\ v_{i2} \\ v_{i3} \end{bmatrix}} = \underset{(3\cdot k)\times 1}{\begin{bmatrix} w_{i1}v_{i1} \\ w_{i2}v_{i2} \\ w_{i3}v_{i3} \end{bmatrix}}.$$

The reason why we use w_i, not q_i, is due to the redundancy of the moment conditions: the 1 in c_i renders $E(v_{it})=0\ \forall\ t$, and thus the time dummies do not give any new moment conditions. We need at least as many instruments as the number of the parameters, which requires

(3.10) $$T\cdot k \geq k + (T-1).$$

This is satisfied so long as $T\geq 2$.

PRE gives different numbers of moments for each t. For T=3,

$E(v_{i1}w_{i1}) = 0 \quad (t=1),$

$E(v_{i2}w_{i1}) = E(v_{i2}w_{i2}) = 0 \quad (t=2),$

$$E(v_{i3}w_{i1}) = E(v_{i3}w_{i2}) = E(v_{i3}w_{i3}) = 0 \quad (t=3).$$

Since c_i appears in all w_{it}, there are redundant moment conditions. Let

$$z_i = \mathrm{diag}\{w_{i1},\ (x_{i1}',w_{i2}')',\ (x_{i1}',x_{i2}',w_{i3}')'\} \tag{3.11}$$

to avoid the redundancy. This renders

$$z_i v_i = \begin{bmatrix} w_{i1} & 0 & 0 \\ 0 & x_{i1} & 0 \\ 0 & w_{i2} & 0 \\ 0 & 0 & x_{i1} \\ 0 & 0 & x_{i2} \\ 0 & 0 & w_{i3} \end{bmatrix} \cdot \begin{bmatrix} v_{i1} \\ v_{i2} \\ v_{i3} \end{bmatrix} = \begin{bmatrix} w_{i1}v_{i1} \\ x_{i1}v_{i2} \\ w_{i2}v_{i2} \\ x_{i1}v_{i3} \\ x_{i2}v_{i3} \\ w_{i3}v_{i3} \end{bmatrix}. \tag{3.12}$$

$[\{k_x\cdot(1+2+3)\} + k_c\cdot 3] \times 3$

With T large, the IVE with this z_i can be cumbersome computationally. The condition analogous to (3.10) is

$$k_x\cdot T(T+1)/2 \ + \ k_c\cdot T \ \ge \ k + (T-1) \tag{3.13}$$

which is satisfied so long as $T\ge 2$.

EXO gives even more moments than PRE, but its analytic expression is simpler using the Kronecker product: with T=3,

$$\begin{aligned} z_i &= \mathrm{diag}\{\ (x_{i1}',x_{i2}',x_{i3}',c_i')',\ (x_{i1}',x_{i2}',x_{i3}',c_i')', \\ &\qquad (x_{i1}',x_{i2}',x_{i3}',c_i')'\ \} \\ &= I_3\otimes(x_{i1}',x_{i2}',x_{i3}',c_i')'. \end{aligned} \tag{3.14}$$

$\{3\cdot(k_x\cdot 3 + k_c)\} \times 3$

This kind of instruments can be seen in GMM for cross-section simultaneous equations where all equations share the same instruments. The condition analogous to (3.10) and (3.13) is

$$k_x\cdot T^2 + k_c\cdot T \ge k + (T-1) \tag{3.15}$$

which is satisfied if $T\ge 2$. As the IVE for PRE, the IVE for EXO can be computationally quite burdensome.

For some models, some variables in w_{it} may be correlated with v_{it} at all leads and lags, or in a limited way (e.g., contemporaneously). Then, we need some straightforward adjustments in above instrument matrices; this will be seen in examples of Section 5.

4. Fixed Effect and Error-Differencing Idea.

In this section, we review error-differencing ideas for fixed-effect models. We will examine the best known example (the "demeaning" estimator) in detail first, which is also called "the" fixed-effect estimator, "within-group" or "covariance" estimator. Also reviewed is the linear-projection idea for fixed-effect.

To describe the demeaning estimator DME, define a matrix purging δ_i out by demeaning:

$$Q_T \equiv I_T - 1_T 1_T'/T$$

which is idempotent and symmetric. Transform (2.2) into

$$(4.1)\qquad Q_T y_i = Q_T m\tau + Q_T(1_T' \otimes c_i)'\alpha + Q_T x_i'\beta + Q_T(1_T\delta_i) + Q_T u_i$$

where

$$Q_T y_i = (y_{i1}-\bar{y}_{i.} \ldots y_{iT}-\bar{y}_{i.})', \qquad \bar{y}_{i.} \equiv (1/T)\Sigma_t y_{it}.$$

Then due to $Q_T c_i=0$ and $Q_T\delta_i=0$, the second and fourth terms in the right-hand side (i.e., all time-invariant variables) disappear to leave

$$(4.2)\qquad Q_T y_i = Q_T m\cdot\tau + Q_T x_i'\beta + Q_T u_i.$$

Now define

$$(4.3)\qquad \underset{T\times(T-1+k_x)}{w_i^{*\prime}} \equiv (m,\ x_i'), \qquad \gamma^* \equiv (\tau',\beta')',$$

and apply LSE to (4.2) under $E(w_i^* Q_T u_i)=0$ to get DME:

$$(4.4)\qquad g_{dme} = (\Sigma_i w_i^* Q_T w_i^{*\prime})^{-1}\cdot\Sigma_i w_i^* Q_T y_i,$$

$$\sqrt{N}(g_{dme}-\gamma^*) =^d N(0,\ \Psi^{-1}\cdot(1/N)\Sigma_i(w_i^* Q_T u_i u_i' Q_T w_i^{*\prime})\cdot\Psi^{-1}),$$

where $\Psi \equiv (1/N)\Sigma_i w_i^* Q_T w_i^{*\prime}$. DME is the LSE on the mean-differenced variables. DME appears e.g., in Anderson and Hsiao (1981,p.599).

Turning to other error-differencing ideas related to IVE, Anderson and Hsiao (1981) introduce a dynamic model with $w_{it}=y_{i,t-1}$ to devise an IVE after removing δ_i by first-differencing $(y_{it}-y_{i,t-1})$; they propose to use $\Delta y_{i,t-2}$ or $y_{i,t-2}$ as an instrument in the section 8 of their paper. They also generalize the IVE further in Anderson and Hsiao (1982, p.78-81) where w_{it} has regressors other than $y_{i,t-1}$.

Holtz-Eakin et al. (1988) show bivariate vector auto regression in panel data by including many lagged dependent variables. They allow δ_i to have a time-varying coefficient Φ_t; $\Phi_t\delta_i$ is then removed together through a transformation (each "wave" (cross-section) is multiplied by Φ_t^{-1} and then first-differenced), to which an IVE with PRE type moment conditions is applied. Allowing Φ_t however complicates identification of the model parameters, for the unknown Φ_t is multiplied into them.

Arellano and Bond (1991) remove δ_i by first-differencing and do GMM using only linear moment conditions that follow from zero serial-correlation in u_{it}. For the dynamic model with $y_{i,t-1}$ in w_{it}, this means that the instruments include more lagged dependent variables than just $y_{i,t-2}$ as in Anderson and Hsiao (1981). Zero serial-correlation can occur in vector-autoregression models and rational expectation models if all lagged variables are included in the information set at t-1; the latter models in fact yield conditional orthogonality which is stronger than unconditional orthogonality. Ahn and Schmidt (1995) propose using nonlinear moment condition as well which are obtained by either restricting error term covariance structure beyond zero serial-correlation or imposing certain stationarity.

Keane and Runkle (1992) do "forward-differencing" $(y_{it}-y_{i,t+1})$ to remove δ_i and then do "forward-filtering", which is a WLS using the upper-triangular Cholesky factor of the variance matrix of $v_{it}-v_{i,t-1}$. The upper-triangular matrix is used to preserve orthogonality of PRE type instruments while making $v_{it}-v_{i,t-1}$ serially uncorrelated; it removes serial correlation with linear combinations of the current and future errors rather than the current and past errors. Using the Cholesky factor forces Keane and Runkle (1992) to imposing second order moment conditions on the error terms.

The idea of forward-filtering first appeared in Hayashi and Sims (1983). They proposed it for time-series data for the same reason of preserving orthogonality of predetermined instruments in rational expectation models with serially correlated errors, which can occur if some recent lagged errors are not included in the information set when the expectation was formed. Arellano and Bover (1995) propose a related differencing idea of subtracting the mean of $y_{i\tau}$'s, $\tau=t,t+1,\ldots T$, from y_{it}, calling it "Helmert" transformation.

Arellano and Bover (1995) transforms v_i into (when T=3)

$$(v_{i2}-v_{1i},v_{i3}-v_{i2},(1/T)\Sigma_t v_{it})' = (u_{i2}-u_{1i},u_{i3}-u_{i2},\delta_i+(1/T)\Sigma_t u_{it})',$$

and classify instruments appropriate for each component of this transformed error vector to do IVE. Arellano and Bover (1995) show that there are other ways to transform v_i one-to-one to, say ν_i, such that the first T-1 components of ν_i are free of δ_i, but that the same estimator is obtained regardless of the form of the transformation.

There are many consistent estimators for fixed-effect models in the literature. Then the choice will depend on computational ease and efficiency. Error-differencing, compared with regressor-differencing, may make computation easier by removing redundant moment conditions. But efficiency-wise, in general, there seems nothing to gain (Schmidt et al. (1992)) or lose (Chamberlain (1992)) by error-differencing, if all relevant moment conditions are used and GMM is employed for regressor-differencing approaches.

The main advantage of error-differencing idea is that, since δ_i is purged out mechanically, we can allow any type of relationship between δ_i and the regressors. The main disadvantage is that we cannot estimate the coefficients of the time-invariants that are also removed along with δ_i. This disadvantage is overcome by regressor-differencing that keeps error terms intact but uses transformed regressors as instruments.

Other than error- and regressor- differencing, there is yet another idea to deal with fixed-effect (thus overall, three approaches for fixed-effect). As a prelude for the third idea, define the linear projection of δ on a vector π whose first component is 1 as

$$E(\delta\pi')\cdot\{E(\pi\pi')\}^{-1}\cdot\pi = \vartheta'\cdot\pi \tag{4.5}$$

where $\vartheta\equiv\{E(\pi\pi')\}^{-1}\cdot E(\pi\delta')$ is the linear projection coefficient; (4.5) is valid even if δ is a vector not a scalar. It is easily seen that $E\{(\delta-\vartheta'\pi)\cdot\pi'\}=0$; i.e, the residual $\delta-\vartheta'\pi$ is uncorrelated with π by construction. Other than $E(\delta\delta')<\infty$ and $E(\pi\pi')<\infty$, the decomposition

$$\delta = \vartheta'\pi + (\delta-\vartheta'\pi) \tag{4.6}$$

is valid for any δ and π.

Following (4.6), Chamberlain (1982,1984) rewrites δ_i as

$$\{c_i'\psi_o+\Sigma_{\tau=1}^{T}x_{i\tau}'\psi_\tau\} + [\delta_i - c_i'\psi_o-\Sigma_{\tau=1}^{T}x_{i\tau}'\psi_\tau] \tag{4.7}$$

where δ_i is projected on $(c_i, x_{i1}'\ldots x_{iT}')'$, and $(\psi_o', \psi_1'\ldots.\psi_T')'$ is the projection coefficient. With (4.7), rewrite (1.4) as

$$y_{it} = \tau_t + c_i'(\alpha+\psi_o) + x_{it}'(\beta+\psi_t) + \Sigma_{\tau\neq t}x_{i\tau}'\psi_\tau \tag{4.8}$$

$$+ [\delta_i - c_i'\psi_o - \Sigma_{\tau=1}^{T} x_{i\tau}'\psi_\tau] + u_{it},$$

where $[\cdot]+u_{it}$ is the new error term, and $\alpha+\psi_o$, $\beta+\psi_t$ and ψ_τ are to be estimated.

The advantage of this approach is that the linear projection does not require any restriction (other than the existence of the second moments), differently from assuming that $E(\delta|c_i, x_{i1}\ldots x_{iT})$ is a known function of c_i and $x_{i1}\ldots x_{iT}$. The disadvantage is identification problems: α is not identified due to ψ_o, and β is not identified immediately (it takes another stage of estimation).

5. Examples of Instrument Matrices.

In this section, we show how to classify regressors as in (1.4) and set up z_i in practice through an example. Recall the example in Section 1 where y_{it} is the annual wage of married men of age 40-60, and

c_i: 1, RC (race), ED (education in years),

x_{it}: WH (working hours), UR (local unemployment rate), SF (1 if self-employed and 0 otherwise),

δ_i: ability, IQ, or productivity;

δ_i may be observable to the employer of the individual i but unobservable to econometricians. Comparing $v_{it}=\delta_i+u_{it}$ with c_i and x_{it}, suppose that we assume the followings:

(5.1) (i) RC_i is not correlated with v_{it};

(ii) ED_i is correlated with v_{it} only through δ_i;

(iii) WH_{it} is correlated with v_{it} only through u_{it} due to the simultaneity with wage;

(iv) UR_{it} is not correlated with v_{it};

(v) SF_{it} is correlated with v_{it} only through δ_i.

Under (5.1), we can use at least 1, RC_i and UR_{it}, t=1...T, as instruments. Also if $COR(WH_{i,t-j}, u_{it})=0$, for some j>0, then the lagged WH may be used as instruments. Furthermore, if we can extract the part of SF_{it} uncorrelated with δ_i, the part and its lagged variables may be used as instruments. For instance, suppose

(5.2) $$E(\delta_i \cdot SF_{it}) \text{ is not a function of } t.$$

Then SF_{it} can be rewritten as (a linear projection idea)

(5.3) $$SF_{it} = \phi_i + SF_{it} - \phi_i \equiv \phi_i + \lambda_{it}$$

where

$$\phi_i \equiv \{E(\delta_i{}^2)\}^{-1} \cdot E(\delta_i SF_{it}) \cdot \delta_i \quad \text{and} \quad \lambda_i \equiv SF_{it} - \phi_i;$$

λ_i is uncorrelated with δ_i by construction. Hence, either of the following (and their lagged versions) can be used as instruments:

(5.4) $$SF_{it} - SF_{i,t-1} = \lambda_{it} - \lambda_{i,t-1},$$

(5.5) $$SF_{it} - (1/T)\Sigma_t SF_{it} = \lambda_{it} - (1/T)\Sigma_t \lambda_{it}.$$

After the regressors are classified as above, suppose we take a simple-minded approach of not using any variable correlated with v_{it}. Then only 1, RC_i and UR_{it} can be used as instruments. Since UR_{it} is time-variant, we can impose various moment conditions on UR_{it}. In the rest of this section, we will set T=3.

Omitting SUM that does not give enough moment conditions, suppose we use CON type conditions

(5.6) $$E(v_{it})=0 \ \forall \ t, \qquad E(RC_i v_{it}) = 0 \ \forall \ t, \qquad E(UR_{it} v_{it}) = 0 \ \forall \ t.$$

Then z_i' becomes

(5.7) $$\begin{bmatrix} 1\ 0\ 0 & RC_i & 0 & 0 & UR_{i1} & 0 & 0 \\ 0\ 1\ 0 & 0 & RC_i & 0 & 0 & UR_{i2} & 0 \\ 0\ 0\ 1 & 0 & 0 & RC_i & 0 & 0 & UR_{i3} \end{bmatrix}$$

where both RC_i and UR_{it} give three instruments. This is not using the time-variant UR_{it} well, as can be seen in the following.

Suppose we use PRE type conditions

(5.8) $E(v_{it})=0\ \forall\ t,\quad E(RC_i v_{it}) = 0\ \forall\ t,\quad E(UR_{is}v_{it}) = 0\ \forall\ s\le t.$

Then z_i' should be

(5.9) $$\begin{bmatrix} 1\ 0\ 0 & RC_i & 0 & 0 & UR_{i1} & 0 & 0 & 0 & 0 & 0 \\ 0\ 1\ 0 & 0 & RC_i & 0 & 0 & UR_{i1} & 0 & UR_{i2} & 0 & 0 \\ 0\ 0\ 1 & 0 & 0 & RC_i & 0 & 0 & UR_{i1} & 0 & UR_{i2} & UR_{i3} \end{bmatrix}$$

which has more instruments than (5.7); compare the columns with UR's in (5.7) and (5.9).

Suppose we use EXO type conditions

(5.10) $E(v_{it})=0\ \forall t,\quad E(RC_i v_{it}) = 0\ \forall\ t,\quad E(UR_{is}v_{it}) = 0\ \forall\ s,t.$

Then z_i' is

(5.11) $(1\ RC_i\ UR_{i1}\ UR_{i2}\ UR_{i3})\otimes I_3 = (1\ RC_i\ UR_i')\otimes I_3$

where $UR_i\equiv(UR_{i1},UR_{i2},UR_{i3})'$; this renders the most instruments.

Now consider using variables correlated with v_{it}, in addition to 1, RC_i and UR_{it}. In this case, basically, we should list all the moment conditions for each v_{it}, t=1...T. Define

$$\Delta SF_{it} \equiv SF_{it} - SF_{i,t-1},\quad \Delta SF_i\equiv(\Delta SF_{i2},\ \Delta SF_{i3})'.$$

Omitting now the CON case which is easy, suppose that we use the following PRE type conditions (recall (5.1)):

(5.12) v_{i1} is orthogonal to 1, RC_i, UR_{i1};

v_{i2} is orthogonal to 1, RC_i, UR_{i1}, UR_{i2}, WH_{i1}, ΔSF_{i2};

v_{i3} is orthogonal to 1, RC_i, UR_{i1}, UR_{i2}, UR_{i3}, WH_{i1}, WH_{i2},

ΔSF_{i2}, ΔSF_{i3}.

Then z_i' is

(5.13)

$$\begin{bmatrix} 1 & RC_i & UR_{i1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & RC_i & UR_{i1} & UR_{i2} & WH_{i1} & \Delta SF_{i2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & RC_i & UR_i' & WH_{i1} & WH_{i2} & \Delta SF_i' \end{bmatrix}$$

where z_i is now written block-diagonally for v_{it}, t=1,2,3.

If we choose to use EXO type conditions for UR_{it}, then z_i' becomes

(5.14)
$$\begin{bmatrix} & 0 & 0 & 0 & 0 & 0 \\ (1, RC_i, UR_i')\otimes I_3 & WH_{i1} & \Delta SF_{i2} & 0 & 0 & 0 \\ & 0 & 0 & WH_{i1} & WH_{i2} & \Delta SF_i' \end{bmatrix}.$$

Compare this with (5.11).

6. Two Special Variables: Age and Experience.

Certain variables in panel data require some care in their use. One example is time dummies and intercept which were already discussed. In this section, we examine age and (job) experience. The fact that these variables require attention can be easily seen in some papers mentioning removing some of these variables to avoid singular matrices. Those remarks are written usually in piecemeal fashion specific to their context; here we explicitly show how the singularity come about through redundant moment conditions.

Consider AGE_{it} which is usually included in the wage example. AGE_{it}

is time-variant, but in a very limited way. It can be rewritten as

$$(6.1) \qquad AGE_{it} = AGE_{i1} + t-1 = AGE_{i0} + t$$

where AGE_{i1} is the age of the individual i at t=1, and $AGE_{i0}=AGE_{i1}-1$.

Examining moment conditions between AGE_{it} and v_{it}, first, suppose we use CON to get

$$(6.2) \qquad E(AGE_{it} \cdot v_{it}) = E(AGE_{i0}v_{it} + t \cdot v_{it}) = 0 \quad \forall\ t.$$

Since $E(tv_{it}) = t \cdot E(v_{it}) = 0$ was already imposed using 1, only $E(AGE_{i0}v_{it})=0$ is new. Second, consider PRE:

$$(6.3) \qquad E(AGE_{is} \cdot v_{it}) = E(AGE_{i0}v_{it} + s \cdot v_{it}) = 0 \quad \forall\ s \le t.$$

Again, only $E(AGE_{i0}v_{it})=0$ is new. For EXO, it is analogous. In short, so long as we have I_T (or its equivalents) in the instrument set, use only AGE_{i1} as a time-invariant instrument.

Often AGE^2 is used along with AGE. From (6.1),

$$(6.4) \qquad AGE_{it}^2 = AGE_{i0}^2 + 2t \cdot AGE_{i0} + t^2$$

which yields

$$(6.5) \qquad E(AGE_{it}^2 \cdot v_{it}) = E\{(AGE_{i0}^2 + 2t \cdot AGE_{i0} + t^2) \cdot v_{it}\} = 0.$$

As in AGE, so long as I_T is in the instrument set, only $E(AGE_{i0} \cdot v_{it})=0$ and $E(AGE_{i0}^2 \cdot v_{it})=0$ are new moment conditions. Thus with AGE and AGE^2, we get only two more time-invariant instruments AGE_{i1} and AGE_{i1}^2. For instance, for EXO in the wage example, expand z_i' in (5.11) into

$$(6.6) \qquad (1\ RC_i\ UR_i'\ \ AGE_{i1}\ AGE_{i1}^2) \otimes I_3.$$

Another notable point for AGE occurs in first-differenced models. Suppose we first-difference (1.4) to get

$$(6.7) \qquad \Delta y_{it} = \Delta\tau_t + \Delta x_{it}'\beta + \Delta u_{it}, \qquad t=2...T.$$

Since $\Delta AGE_{it}=1$, the column for AGE becomes 1_{T-1}. If we include 1_{T-1} in (6.7) to estimate the intercept, we will end up with two 1_{T-1} columns.

Using time dummies and AGE in the first-differenced model creates a problem. Suppose T=3 and we have (I_3, AGE_{it}). Then taking the first difference, we get

$$(6.8)\qquad \begin{bmatrix} 1 & 0 & 0 & AGE_{i1} \\ 0 & 1 & 0 & AGE_{i2} \\ 0 & 0 & 1 & AGE_{i3} \end{bmatrix} \rightarrow \begin{bmatrix} -1 & 1 & 0 & 1 \\ 0 & -1 & 1 & 1 \end{bmatrix}.$$

Thus we need to drop any two out of these four columns, say the middle two. For a generic T, we would get a $(T-1)\times(T+1)$ matrix, from which two columns should be removed.

Besides AGE, often we have a job experience JE_{it} in the regressor. With no unemployment spell, JE_{it} is little different from AGE_{it}, for

$$(6.9)\qquad JE_{it} = JE_{i1} + (t-1) = JE_{i0} + t.$$

So above cautions on AGE_{it} also hold for JE_{it} with no unemployment spell. If there are unemployment spells but if T is small and most people do not change their employment status, then still there can be a problem of nearly singular matrices; regarding JE_{it} as (6.9) and thus using only JE_{i1} is a solution to avoid the near singularity.

7. Regressor-Differencing with SUM Type Conditions and Efficiency Gain.

In the example (5.1), one of the main interests is on education ED, the question called "returns to schooling". Since ED_i is likely to be correlated with δ_i, a fixed-effect model estimator is called for. But error-differencing will not do, for ED_i is removed along with δ_i; it is necessary to estimate the model without purging out all time-invariants. This can be done by IVE with regressor-differencing, the literature of which, however, has been limited more or less to SUM type conditions under either homoskedasticity or heteroskedasticity of known form. In the following, we will set T=3, and review this part of the literature, and then discuss an efficiency gain issue using more instruments.

Hausman and Taylor (1981) note that the time-invariant part of the time-variant variables may be used as instruments for the endogenous time-invariant variables; in the wage example, $UR_{i.} \equiv (1/T)\Sigma_t UR_{it}$ can be used as an instrument for ED_i. To simplify exposition, recall (5.1) and consider only 1, RC_i and UR_{it} as the source for instrument. Under SUM, the instrument matrix z_i' would have looked like

(7.1)
$$\begin{bmatrix} 0 & 0 & 1 & RC_i & UR_{i1} \\ 1 & 0 & 1 & RC_i & UR_{i2} \\ 0 & 1 & 1 & RC_i & UR_{i3} \end{bmatrix}$$

which would be insufficient for the model with ED_i in. According to Hausman and Taylor (1981), z_i' should be augmented as

(7.2)
$$\begin{bmatrix} 0 & 0 & 1 & RC_i & UR_{i1} & UR_{i.} \\ 1 & 0 & 1 & RC_i & UR_{i2} & UR_{i.} \\ 0 & 1 & 1 & RC_i & UR_{i3} & UR_{i.} \end{bmatrix}$$

where $UR_{i.}$ may be correlated with ED_i; highly educated people may live in low unemployment area.

Amemiya and MaCurdy (1986) suggest including the time-invariant variables generated from the exogenous time-variant variables. In the wage example, z_i' should have three new columns in the place of $UR_{i.}$:

(7.3)
$$\begin{bmatrix} 0 & 0 & 1 & RC_i & UR_{i1} & UR_{i1} & UR_{i2} & UR_{i3} \\ 1 & 0 & 1 & RC_i & UR_{i2} & UR_{i1} & UR_{i2} & UR_{i3} \\ 0 & 1 & 1 & RC_i & UR_{i3} & UR_{i1} & UR_{i2} & UR_{i3} \end{bmatrix}.$$

Breusch et al.(1989) further extend the instrument matrix by including the time-invariant variables generated from the time-variant variables correlated with δ_i under the assumption (5.2). In the wage example, we can use the self employment dummy variable SF_{it}. The instrument matrix z_i' becomes

$$(7.4) \qquad \begin{bmatrix} 0 & 0 & 1 & RC_i & UR_{i1} & UR_{i1} & UR_{i2} & UR_{i3} & \Delta SF_{i2} & \Delta SF_{i3} \\ 1 & 0 & 1 & RC_i & UR_{i2} & UR_{i1} & UR_{i2} & UR_{i3} & \Delta SF_{i2} & \Delta SF_{i3} \\ 0 & 1 & 1 & RC_i & UR_{i3} & UR_{i1} & UR_{i2} & UR_{i3} & \Delta SF_{i2} & \Delta SF_{i3} \end{bmatrix}.$$

In (7.1) to (7.4), we have increasingly more time-invariant instruments generated by time-variant variables. But comparing (5.7), (5.9) and (5.11) with (7.1) to (7.4), the latter use only SUM type conditions while the former use far more. This has some interesting implication on efficiency gain in using more instrument as shown in the rest of this section.

After above three papers on IVE with increasingly more instruments, there came up a question on the efficiency gain. Cornwell and Ruppert (1988) applied the IVE's in the three papers to the returns-to-schooling problem, and noted that the efficiency gain was limited to the coefficient of the endogenous time-invariant regressor ED_i. This was plausible, for the new instruments are all time-invariant. Baltagi and Khanti-Akom (1990) used (almost) the same data set as Cornwell and Ruppert (1988) used, but with different classifications of the instruments. They observed that efficiency gain was not limited to the endogenous time-invariant regressor and that the efficiency gain is much smaller than that in Cornwell and Ruppert (1988). We will show there should be no efficiency gain for the endogenous time-variant variables so long as SUM is used; we need at least CON.

Let T=2 and consider a time-variant 2×1 random vector

$m_i \equiv (m_{i1}, m_{i2})'$. Also consider two time-invariant random variables d_i and e_i (both scalar). Consider approximating m_i by d_i and e_i. There are a couple of different ways. One way is to choose λ_d and λ_e in

$$(7.5) \qquad \begin{bmatrix} m_{i1} \\ m_{i2} \end{bmatrix} - \lambda_d \cdot \begin{bmatrix} d_i \\ d_i \end{bmatrix} - \lambda_e \cdot \begin{bmatrix} e_i \\ e_i \end{bmatrix},$$

such that this deviation is small in a sense. In this case, both m_{i1} and m_{i2} are fitted by $\lambda_d d_i + \lambda_e e_i$ which is time-invariant. Another way is to choose λ_{d1}, λ_{d2}, λ_{e1} and λ_{e2} in

$$(7.6) \qquad \begin{bmatrix} m_{i1} \\ m_{i2} \end{bmatrix} - \lambda_{d1} \begin{bmatrix} d_i \\ 0 \end{bmatrix} - \lambda_{d2} \begin{bmatrix} 0 \\ d_i \end{bmatrix} - \lambda_{e1} \begin{bmatrix} e_i \\ 0 \end{bmatrix} - \lambda_{e2} \begin{bmatrix} 0 \\ e_i \end{bmatrix},$$

such that this is small in a sense. Here m_{i1} is fitted by $\lambda_{d1} d_i + \lambda_{e1} e_i$ and m_{i2} is fitted by $\lambda_{d2} d_i$ and $\lambda_{e2} e_i$: d_i and e_i can generate a time-variant feature, so long as d_i and e_i are not independent of m_{i1} and m_{i2}. (7.5) imposes $\lambda_{d1} = \lambda_{d2}$ and $\lambda_{e1} = \lambda_{e2}$ in comparison with (7.6).

Recall the wage example where WH_{it} (working hours) is the endogenous time-variant variable. Let $d_i = UR_{i1}$ and $e_i = UR_{i2}$. Then it is possible that WH_{i1} is explained by UR_{i1}, while WH_{i2} is explained by UR_{i1} (lagged unemployment rate) and UR_{i2} (current unemployment rate). Using (7.3) or (7.4) leads to (7.5), while using (5.9) or (5.13) leads to (7.6). Hence, with the moment conditions as in CON, PRE, and EXO, there can be efficiency gain for the endogenous time-variant regressors as well as for the endogenous time-invariant regressors.

Baltagi and Khanti-Akom (1990) use a SUM type moment condition where there should be no efficiency gain for the endogenous time-variant regressors. Nevertheless, they noted an efficiency gain in one endogenous time-variant variable, "job experience" (JE_{it}). As discussed in (6.9) and also noted in their paper, barring unemployment spells, (7.7) $$JE_{it} = JE_{i1} + t-1.$$

With time dummies (I_T) in, the only part of JE_{it} left to be explained by the time-invariant instruments is the time-invariant JE_{i1}. So it is not surprising to see an efficiency gain in JE_{it}.

8. MDE When Number of Instruments Varies across Time.

When the number of instruments varies across time, putting together all the different moment conditions for the different periods is practically rather cumbersome. In this section, we will estimate each wave separately, and combine the estimates later with a MDE using cross-equation restrictions. Using the MDE, we can also test the constancy of the parameters across t. If the test rejects the constancy, then (1.4) should be extended to allow time-varying parameters; this is done in (8.17) later. It'll be seen there that our IVE and GMM are still applicable to (8.17), if at least CON type conditions are used.

Consider the period t with k_t-many instruments. Define the following matrices for the tth wave:

(8.1) $$Y_t = W_t\gamma_t + V_t$$

where

(8.2) $$\underset{N\times 1}{Y_t}, \quad \underset{N\times k}{W_t}, \quad \underset{N\times 1}{V_t}, \quad \underset{N\times k_t}{Z_t};$$

Z_t is the $N\times k_t$ instrument matrix. The IVE $g_{t,ive}$ for the parameter γ_t is

(8.3) $$g_{t,ive} \equiv \{W_t'Z_t(Z_t'Z_t)^{-1}Z_t'W_t\}^{-1}W_t'Z_t(Z_t'Z_t)^{-1}Z_t'Y_t.$$

Also the GMM g_t is

(8.4) $$g_t \equiv \{W_t'Z_t(Z_t'D_tZ_t)^{-1}Z_t'W_t\}^{-1}W_t'Z_t(Z_t'D_tZ_t)^{-1}Z_t'Y_t$$

where D_t is $\mathrm{diag}(\hat{V}_t)$ and $\hat{V}_t \equiv Y_t - W_t g_{t,ive}$. The IVE (8.3) can be seen also in Holtz-Eakin et al.(1988, footnote 12); they use it not for the MDE but to easily get $\hat{v}_i$ in (3.6).

Using the vector notation, instead of the matrix,

$$(8.5) \qquad g_{t,ive}-\gamma_t = \{\Sigma_i w_{it}z_{it}'/N\cdot(\Sigma_i z_{it}z_{it}'/N)^{-1}\cdot\Sigma_i z_{it}w_{it}'/N\}^{-1}$$
$$\cdot\Sigma_i w_{it}z_{it}'/N\cdot(\Sigma_i z_{it}z_{it}'/N)^{-1}\Sigma_i z_{it}v_{it}/N;$$

$$(8.6) \qquad g_t-\gamma_t = \{\Sigma_i w_{it}z_{it}'/N\cdot(\Sigma_i z_{it}z_{it}'\hat{v}_{it}^{\,2}/N)^{-1}\cdot\Sigma_i z_{it}w_{it}'/N\}^{-1}$$
$$\cdot\Sigma_i w_{it}z_{it}'/N\cdot(\Sigma_i z_{it}z_{it}'\hat{v}_{it}^{\,2}/N)^{-1}\Sigma_i z_{it}v_{it}/N$$

where $\hat{v}_{it}$'s are the components of $\hat{V}_t$.

It is possible that there may not be enough instruments when t is close to 1. Then those periods cannot be used for the MDE, which will entail certain efficiency loss. If T is small and this problem occurs, then it will be better to use the GMM by stacking up all moment conditions (as in (5.13)) rather than the MDE.

Suppose T=3 and we have the estimates g_1, g_2 and g_3 for γ_1, γ_2 and γ_3 respectively; let $g_N \equiv (g_1', g_2', g_3')'$. The restriction used for MDE is that the slope coefficients μ_t in γ_t, t=1,2,3, are the same, while the intercepts showing the time trend are allowed to be different. Denoting the $(k-1)\times 1$ common slope parameter vector as μ, these restrictions are

$$(8.7) \qquad \begin{bmatrix} \tau_1 \\ \mu_1 \\ \tau_2 \\ \mu_2 \\ \tau_3 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{k-1} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & I_{k-1} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & I_{k-1} \end{bmatrix} \cdot \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \mu \end{bmatrix}.$$

Denote the matrix with 1 and I_{k-1} as R, and define $\gamma_o \equiv (\tau_1, \tau_2, \tau_3, \mu')'$.

Subtract g_N from both sides of (8.7) to get

$$(8.8)\qquad \begin{bmatrix} g_1-\gamma_1 \\ g_2-\gamma_2 \\ g_3-\gamma_3 \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix} - R\cdot\gamma_o.$$

The idea of MDE is estimating γ_o by minimizing

$$(8.9)\qquad (g_N-R\cdot g_o)'W^{-1}(g_N-R\cdot g_o)$$

wrt g_o where W is a positive definite matrix. The efficient choice of W is $V[g_N-R\gamma_o]=V[g_N-\gamma]$ owing to (8.8). To get $V[g_N-\gamma]$, define A_t and ψ_{it} such that

$$(8.10)\qquad g_t-\gamma_t \equiv (1/N)\Sigma_i A_t z_{it} v_{it} \equiv (1/N)\Sigma_i \psi_{it};$$

i.e, A_t is the $k\times k_t$ matrix in front of $\Sigma_i z_{it} v_{it}/N$ in (8.6). Finally define $\psi_i \equiv (\psi_{i1}', \psi_{i2}', \psi_{i3}')'$ to get

$$(8.11)\qquad V[g_N-\gamma] =^p (1/N)\Sigma_i \psi_i \psi_i'/N \equiv W_N.$$

The solution g_{mde} for (8.9) is

$$(8.12)\qquad g_{mde} = (R'W_N{}^{-1}R)^{-1}\cdot R'W_N{}^{-1}g_N;$$

$$\sqrt{N}(g_{mde}-\gamma) =^d N(0,\ (R'W_N{}^{-1}R)^{-1}).$$

Analogously to GMM over-identifying restriction tests, we get

$$(8.13)\qquad (g_N-R\cdot g_{mde})'W_N{}^{-1}(g_N-R\cdot g_{mde}) =^d \chi^2_{2(k-1)}$$

under the null hypothesis (8.7). Alternatively, we may use the following idea to test (8.7).

At each t, we obtain an estimate m_t for μ_t which is supposed to be more efficient than m_{t-1} which is in turn supposed to be more efficient than m_{t-2} and so on. Then we can devise a simple test for the change in the slopes:

$$(8.14)\qquad (m_t-m_{t-1})'\cdot\{V[m_{t-1}-m_t]\}^{-1}\cdot(m_t-m_{t-1}) =^d \chi_{k-1}^2.$$

To estimate $\{V[\cdot]\}^{-1}$, define a scalar ψ_{1it} and a $(k-1)\times1$ vector ψ_{sit} such that $\psi_{it}\equiv(\psi_{1it},\psi_{sit}')'$, t=1,2,3, (in the subscripts, "s" is for the slope). Also define the "effective score" ψ_{sit}^* for the slope as

$$\psi_{sit}^* = \psi_{sit} - \{\Sigma_i\psi_{sit}\psi_{1it}/\Sigma_i\psi_{1it}^2\}\cdot\psi_{1it}.$$

This yields,

$$m_t-\mu_t = (1/N)\Sigma_i\psi_{sit}^*,\quad t=1\ldots T.$$

Then, under $H_o:\mu_{t-1}=\mu_t$,

$$(8.15)\qquad V[m_{t-1}-m_t] =^p (1/N)\Sigma_i(\psi_{sit}^*-\psi_{si,t-1}^*)(\psi_{sit}^*-\psi_{si,t-1}^*)'/N.$$

If we reject (8.7), then we need to allow the slope coefficient to be time-varying by generalizing the model (1.4) into

$$(8.16)\qquad y_{it} = \underset{1\times k_c}{c_i'}\alpha_t + \underset{1\times k_x}{x_{it}'}\beta_t + \delta_i + u_{it} \equiv \underset{1\times k}{w_{it}'}\gamma_t + v_{it}.$$

Since the first component of α_t is the intercept which varies across t, now we do not need any time dummy variable.

Stacking up the model (8.16) for the unit i across t,

$$(8.17)\qquad \underset{T\times1}{y_i} = \underset{T\times(T\cdot k_c)}{(I_T\otimes c_i)'}\alpha + \underset{T\times(T\cdot k_x)}{x_i'}\beta + \underset{T\times1}{1_T}\delta_i + u_i,$$

$$\equiv \underset{T\times(T\cdot k)}{w_i'}\gamma + v_i,\qquad i=1,\ldots,N.$$

where $\alpha\equiv(\alpha_1',\ldots,\alpha_T')'$ is a $(T\cdot k_c)\times1$ vector, $\beta\equiv(\beta_1',\ldots,\beta_T')'$ is a $(T\cdot k_x)\times1$ vector, $x_i\equiv\mathrm{diag}(x_{i1}',\ldots,x_{iT}')$ is a $T\times(T\cdot k_x)$ matrix, and

$$w_i' \equiv [\ (I_T\otimes c_i)'\quad x_i'\ ],\qquad \gamma \equiv (\alpha',\ \beta')'.$$

Applying IVE and GMM to (8.17) is straightforward. The only change needed is that there should be at least $T\cdot k$ many instruments. For this, we should impose at least CON. Since this is barely enough, PRE or EXO may be better in practice for the time-varying coefficient model.

It is possible to generalize (8.16) further by allowing the effect of the unit-specific term δ_i on y_{it} to vary across t as done in Holtz-Eakin et al.(1988). Suppose

$$(8.18) \qquad \underset{T\times 1}{y_i} = \underset{T\times(T\cdot k_c)}{(I_T \otimes c_i)'}\alpha + \underset{T\times(T\cdot k_x)}{x_i'}\beta + \underset{T\times 1}{\Phi}\cdot\delta_i + u_i,$$

where $\Phi=(\Phi_1 \ldots \Phi_T)'$. This however does not change our estimators, since they are based on orthogonality between δ_i and instruments. Φ becomes a trouble only when one tries to either remove or estimate Φ.

9. Conclusions.

In this paper, we reviewed IVE for linear panel data models, using unconditional orthogonality conditions between the error terms and instruments. Since no other restrictions are imposed in the way of conditional moment conditions or distributional assumptions such as stationarity, we were able to allow a fairly general model. Classifying the orthogonality conditions into four categories provided an uniform forum to compare various IVE's in the literature. In addition to the review, we addressed the question of efficiency gain in using more instruments: the gain does not have to be limited to endogenous time-invariants. We also showed how to do minimum distance estimation, and extended our model to test and allow for time-varying parameters.

REFERENCES

Ahn, S.C. and P.Schmidt, 1995, Efficient estimation of models for dynamic panel data, Journal of Econometrics 68, 5-27.

Amemiya,T. and T.E.MaCurdy, 1986, Instrumental variable estimation of an error-components model, Econometrica 54, 869-881.

Anderson,T.W. and C.Hsiao, 1981, Estimation of dynamic models with error components, Journal of American Statistical Association 81, 598-606.

Anderson, T.W. and C.Hsiao, 1982, Formulation and estimation of dynamic models using panel data, Journal of Econometrics 18, 47-82.

Arellano, M and S.Bond, 1991, Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, Review of Economic Studies 58, 277-297.

Arellano, M. and O.Bover, 1995, Another look at the instrumental variable estimation of error-components models, Journal of Econometrics 68, 29-51.

Baltagi, B.H. and S.Khanti-Akom, 1990, On efficient estimation with panel data, Journal of Applied Econometrics 5, 401-406.

Breusch, T.S.,G.E.Mizon and P.Schmidt, 1989, Efficient estimation using panel data, Econometrica 57, 695-700.

Chamberlain, G., 1982, Multivariate regression models for panel data, Journal of Econometrics 18, 5-46.

Chamberlain, G., 1984, Panel data, in Handbook of Econometrics, edited by Z.Griliches and M.Intrilligator, North Holland.

Chamberlain, G., 1987, Asymptotic efficiency in estimation with conditional moment restrictions, Journal of Econometrics 34, 305-334.

Chamberlain, G., 1992, Comment: sequential moment restrictions in panel data, Journal of Business and Economic Statistics 10, 20-26.

Cornwell,C and P.Ruppert, 1988, Efficient estimation with panel data: An empirical comparison of instrumental variables estimators, Journal of Applied Econometrics 3, 149-155.

Hansen, L., 1982, Large sample properties of generalized method of moments estimators, Econometrica 50, 1029-1054.

Hausman, J.A. and W.E.Taylor, 1981, Panel data and unobservable individual effects, Econometrica 49, 1377-1398.

Hayashi, F. and C.Sims, 1983, Nearly efficient estimation of time series models with predetermined but not exogenous instruments, Econometrica 51, 783-792.

Holtz-Eakin, D., W.Newey and H.S.Rosen, 1988, Estimating vector autoregressions with panel data, Econometrica 56, 1371-1395.

Keane, M.P. and D.E.Runkle, 1992, On the estimation of panel-data models with correlation when instruments are not strictly exogenous, Journal of Business and Economic Statistics 20, 1-9.

Lee, L.F., 1992, Amemiya's generalized least squares and tests of over-identification in simultaneous equation models with qualitative or limited dependent variables, Econometric Reviews 11, 319-328.

Lee, M.J., 1996, Methods of moments and semiparametric econometrics for limited dependent variable models, Springer-Verlag.

Newey, W. and D.McFadden, 1994, Large sample estimation and hypothesis testing, in Handbook of Econometric IV, North-Holland.

Robinson, P., 1987, Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form, Econometrica 55, 875-891.

Schmidt, P., S.C.Ahn, and D.Wyhowski, 1992, Comment, Journal of Business and Economic Statistics 10, 10-14.

No.	Author(s)	Title
9575	A.B.T.M. van Schaik and H.L.F. de Groot	Unemployment and Endogenous Growth
9576	A.J.T.M. Weeren, J.M. Schumacher and J.C. Engwerda	Coordination in Continuously Repeated Games
9577	A. van den Nouweland, S. Tijs and M. Wooders	Axiomatizations of Lindahl and Ratio Equilibria in Public Good Economies
9578	Richard F. Hartl and Peter M. Kort	Capital Accumulation of a Firm Facing an Emissions Tax
9579	S. Eijffinger and E. Schaling	Optimal Commitment in an Open Economy: Credibility Vs. Flexibility
9580	Willem J.H. Van Groenendaal	Estimating Net Present Value Variability for Deterministic Models
9581	M. Perry and P.J. Reny	A General Solution to King Solomon's Dilemma
9582	H. Huizinga and S.B. Nielsen	Capital Income and Profits Taxation with Foreign Ownership of Firms
9583	T. Berglund and R. Kabir	What Explains the Difference Between the Futures' Price and its "Fair" Value? Evidence from the European Options Exchange
9584	F. Janssen, T. de Kok and F. van der Duyn Schouten	Approximations for the Delivery Splitting Model
9585	E. Canton	Labour Supply Shocks and Neoclassical Theory
9586	E. Ley, M. Steel	A Model of Management Teams
9587	S. Eijffinger, J. de Haan	The Political Economy of Central Bank Independence
9588	J. Suijs, P. Borm, A. De Waegenaere, S. Tijs	Cooperative games with stochastic payoffs
9589	J. Ziliak and T. Kniesner	Estimating Life-Cycle Labor Tax Effects
9590	M. van de Ven	Public pensions in a Representative Democracy
9591	C. Fernandez, M. Steel	Reference Priors in Non-Normal Location Problems
9592	K. Wärneryd	Demystitifying Rational Expectations Theory through an Economic-Psychological Model
9593	R. Alessie, A. Lusardi, A. Kapteyn	Saving and Wealth Holdings of the Elderly
9594	R. Joosten and D. Talman	A Globally Convergent Price Adjustment Process for Exchange Economies

No.	Author(s)	Title
9595	E. Calvo, J. Lasaga and A. v.d. Nouweland	Probabilistic graphs in cooperative games: a model inspired by voting situations
9596	R. Douven and J. Plasman	Convergence and International Policy Coordination in the EU: a Dynamic Games Approach
9597	H. Uhlig	A Toolkit Analyzing Nonlinear Dynamic Stochastic Models Easily
9598	M. Lettau and T. Van Zandt	Robustness of Adaptive Expectations as an Equilibrium Selection Device
9599	H. Bloemen	The Relation between Wealth and Labour Market Transitions: an Empirical Study for the Netherlands
95100	J. Blanc and L. Lenzini	Analysis of Commmunication Systems with Timed Token Protocols using the Power-Series Algorithm
95101	R. Beetsma and L. Bovenberg	The Interaction of Fiscal and Monetary Policy in a Monetary Union: Balancing Credibility and Flexibility
95102	P. de Bijl	Aftermarkets: The Monopoly Case
95103	F. Kumah	Unanticipated Money and the Demand for Foreign Assets - A Rational Expectations Approach
95104	M. Vázquez-Brage, A. van den Nouweland, I. García-Jurado	Owen's Coalitional Value and Aircraft Landing Fees
95105	Y. Kwan and G. Chow	Estimating Economic Effects of the Great Leap Forward and the Cultural Revolution in China
95106	P. Verheyen	The missing Link in Budget Models of Nonprofit Institutions; Two Practical Dutch Applications
95107	J. Miller	Should we Offer the Unemployment Places on Labour Market Programmes With the Intention That They Reject Them?
95108	C. van Raalte and H. Webers	Statial Competition with Intermediated matching
95109	W. Verkooijen, J. Plasmans and H. Daniels	Long-Run Exchange Rate Determination: A Neutral Network Study
95110	E. van der Heijden, J. Nelissen, J. Potters and H. Verbon	Transfers and Reciprocity in Overlapping-Generations Experiments
95111	A. van den Elzen and D. Talman	An Algorithmic approach towards the Tracing Procedure of Harsanyi and Selten

No.	Author(s)	Title
95112	R. Beetsma and L. Bovenberg	Does Monetary Unification Lead to Excessive Debt Accumulation?
95113	H. Keuzenkamp	Keynes and the Logic of Econometric Method
95114	E. Charlier, B. Melenberg and A. van Soest	Estimation of a Censored Regression Panel Data Model Using Conditional Moment Restrictions Efficiently
95115	N. G. Noorderhaven, B. Nooteboom and H. Berger	Exploring Determinants of Perceived Interfirm Dependence in Industrial Supplier Relations
95116	F. Kleibergen and H. Hoek	Bayesian Analysis of ARMA models using Noninformative Priors
95117	J. Lemmen and S. Eijffinger	The Fundamental Determiniants of Financial Integration in the European Union
95118	B. Meijboom and J. Rongen	Clustering, Logistics, and Spatial Economics
95119	A. de Jong, F. de Roon and C. Veld	An Empirical Analysis of the Hedging Effectiveness of Currency Futures
95120	J. Miller	The Effects of Labour Market Policies When There is a Loss of Skill During Unemployment
95121	S. Eijffinger, M. Hoeberichts and E. Schaling	Optimal Conservativeness in the Rogoff (1985) Model: A Graphical and Closed-Form Solution
95122	W. Ploberger and H. Bierens	Asymptotic Power of the Integrated Conditional Moment Test Against Global and Large Local Alternatives
95123	H. Bierens	Nonparametric Cointegration Analysis
95124	H. Bierens and W. Ploberger	Asymptotic Theory of Integrated Conditional Moment Tests
95125	E. van Damme	Equilibrium Selection in Team Games
95126	J. Potters and F. van Winden	Comparative Statics of a Signaling Game: An Experimental Study
9601	U. Gneezy	Probability Judgements in Multi-Stage Problems: Experimental Evidence of Systematic Biases
9602	C. Fernández and M. Steel	On Bayesian Inference under Sampling from Scale Mixtures of Normals
9603	J. Osiewalski and M. Steel	Numerical Tools for the Bayesian Analysis of Stochastic Frontier Models
9604	J. Potters and J. Wit	Bets and Bids: Favorite-Longshot Bias and Winner's Curse
9605	H. Gremmen and J. Potters	Assessing the Efficacy of Gaming in Economics Educating

No.	Author(s)	Title
9606	J. Potters and F. van Winden	The Performance of Professionals and Students in an Experimantal Study of Lobbying
9607	J.Kleijnen, B. Bettonvil and W. van Groenendaal	Validation of Simulation Models: Regression Analysis Revisited
9608	C. Fershtman and N. Gandal	The Effect of the Arab Boycott on Israel: The Automobile Market
9609	H. Uhlig	Bayesian Vector Autoregressions with Stochastic Volatility
9610	G. Hendrikse	Organizational Change and Vested Interests
9611	F. Janssen, R. Heuts and T. de Kok	On the (R,s,Q) Inventory Model when Demand is Modelled as a compound Bernoulli Process
9612	G. Fiestras-Janeiro P. Borm and F. van Megen	Protective Behaviour in Games
9613	F. van Megen, G. Facchini, P. Borm and S. Tijs	Strong Nash Equilibria and the Potential Maximizer
9614	J. Miller	Do Labour Market Programmes Necessarily Crowd out Regular Employment? - A Matching Model Analysis
9615	H. Huizinga	Unemployment Benefits and Redistributive Taxation in the Presence of Labor Quality Externalities
9616	G. Asheim and M. Dufwenberg	Admissibility and Common Knowlegde
9617	C. Fernández, J. Osiewalski and M. Steel	On the Use of Panel Data in Bayesian Stochastic Frontier Models
9618	H. Huizinga	Intrafirm Information Transfers and Wages
9619	B. Melenberg and B. Werker	On the Pricing of Options in Incomplete Markets
9620	F. Kleibergen	Reduced Rank Regression using Generalized Method of Moments Estimators
9621	F. Janssen, R. Heuts and T. de Kok	The Value of Information in an (R,s,Q) Inventory Model
9622	F. Groot, C. Withagen and A. de Zeeuw	Strong Time-Consistency in the Cartel-Versus-Fringe Model
9623	R. Euwals and A. van Soest	Desired and Actual Labour Supply of Unmarried Men and Women in the Netherlands
9624	M. Khanman, M. Perry and P.J. Reny	An Ex-Post Envy-Free and Efficient Allocation Mechanism: Imperfect Information Without Common Priors

No.	Author(s)	Title
9625	C. Eaves, G. van der Laan D. Talman and Z. Yang	Balanced Simplices on Polytopes
9626	H. Bloemen and A. Kalwij	Female Employment and Timing of Births Decisions: a Multiple State Transition Model
9627	P. Bolton	Strategic Experimentation: a Revision
9628	C. Fershtman and Y. Weiss	Social Rewards, Externalities and Stable Preferences
9629	P.M. Kort, G. Feichtinger R.F. Hartl and J.L. Haunschmied	Optimal Enforcement Policies (Crackdowns) on a Drug Market
9630	C. Fershtman and A. de Zeeuw	Tradeable Emission Permits in Oligopoly
9631	A. Cukierman	The Economics of Central Banking
9632	A. Cukierman	Targeting Monetary Aggregates and Inflation in Europe
9633	F. de Roon, C. Veld and J. Wei	A Study on the Efficiency of the Market for Dutch Long Term Call Options
9634	B. van Aarle, L. Bovenberg M. Raith	Money, Fiscal Deficits and Government Debt in a Monetary Union
9635	G. van der Laan, D. Talman and Z. Yang	Existence of and Equilibrium in a Competitive Economy with Indivisibilities and Money
9636	I. Brouwer, J. van der Put C. Veld	Contrarian Investment Strategies in a European Context
9637	M. Berg, F. van der Duyn Schouten and J. Jansen	Optimal Service Policies to Remote Customers with Delay-Limits
9638	F.C. Drost and C.A.J. Klaassen	Efficient Estimation in Semiparametric GARCH Models
9639	A.G. de Kok and F.B.S.L.P. Janssen	Demand Management in Multi-Stage Distribution Chain
9640	D.A. Freund, T.J. Kniesner A.T. LoSasso	How Managed Care Affects Medicaid Utilization A Synthetic Difference-in-Differences Zero-Inflated Count Model
9641	M.J. Lee	Instrumental Variable Estimation For Linear Panel Data Models